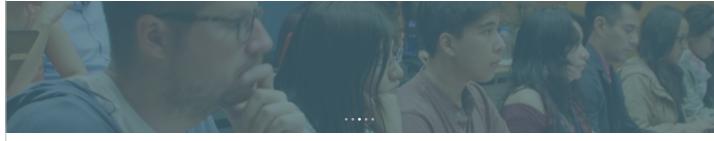


- Bienvenidos al TIB2024-1 Pangenómica y Filogenómica microbiana



- Profesor: **Pablo Vinuesa**, <http://www.ccg.unam.mx/~vinuesa/> @pvinmex
- Ayudantes: **Daniela Hernández y Mauricio Osorio**
 - Acceso al material del curso:
<https://github.com/vinuesa/TIB-filoinfo>
 - Si tienes una laptop con Windows, necesitas instalar mobaXterm. Ver instrucciones in la url arriba indicada

- Bienvenidos al TIB2024-1 Pangenómica y Filogenómica microbiana



17 participantes de Querétaro, Michoacán, CDMX, Hidalgo, Guanajuato y Morelos
 Estudiantes de licenciatura, posgrado e investigadores

Nombre(s)	Apellidos	Instución, Universidad o Empresa	Facultad, Dependencia, Departamento o Ár/Estado	Posición laboral
Ernesto	Garfias Morales	UNAM Campus Juriquilla	LIGH	Querétaro Labmam
Melina	Ortiza	UNAM Campus Morelia	Escuela Nacional de Estudios Superiores	Méjico Pasante de licenciatura
YANIQUE	Tejópila Pineda	UNAM Campus Central	Instituto Politécnico Nacional	Ciudad de México Estudiante de Posgrado
Maria Alessandra	Villaseca BENITEZ	UNAM	Instituto de Investigaciones Biomédicas	Ciudad de México Estudiante de Posgrado
Carlos Adrián	García Ausencio	UNAM	Instituto de Investigaciones Biomédicas	Ciudad de México Estudiante de Posgrado
Alberto	Patricio-Hernández	Universidad Politécnica de Pachuca	Laboratorio de agrobiotecnología	Hidalgo Estudiante de Posgrado
Guillermo de Jesus	Castillo Cortes	UA del Estado de Hidalgo	Escuela Superior de Apan	Hidalgo Profesional
Erick Daniel	Hernández García	UA del Estado de Hidalgo	Escuela Superior de Apan	Hidalgo Estudiante de licenciatura
Aaron	Miranda Neri	UA del Estado de Hidalgo	Escuela Superior de Apan	Hidalgo Estudiante de licenciatura
Karina	Mendoza Gómez	UA del Estado de Hidalgo	Escuela Superior de Apan	Hidalgo Estudiante de licenciatura
Valeria Elena	Bermeo Jaime	UA del Estado de Hidalgo	Escuela Superior de Apan	Hidalgo Estudiante de licenciatura
Dafne Abigail	Guillén Navarro	Instituto Politécnico Nacional	Escuela Nacional de Ciencias Biológicas	Ciudad de México Estudiante de Posgrado
Stephanie	Mata González	UNAM	Facultad de Ciencias	Ciudad de México Estudiante de Posgrado
Laura María	Contreras Alvarado	Hosp. Infantil de México Federico Gómez	Lab. de Investig. en Bacteriología Intestinal	Ciudad de México Estudiante de Posgrado
Zulia Fernandina	Nieves López	IPN-Unidad Irapuato	Departamento de Ingeniería Genética	Guanajuato Estudiante de Posgrado
José Daniel	Chávez González	IPN-Unitad Irapuato	Ingeniería Genética	Guanajuato Estudiante de Posgrado
GUILLERMO	MARQUEZ LICONA	Instituto Politécnico Nacional	Centro de Desarrollo de Productos Bióticos	Morelos Investigador



Introducción a la Filoinformática: Pan-genómica y Filogenómica Microbiana –
 NNB & CCG-UNAM, 22-26 de enero 2024

Pablo Vinuesa (vinuesa@ccg.unam.mx; [@pvinmex](http://www.ccg.unam.mx/~vinuesa/))
 Centro de Ciencias Genómicas, CCG-UNAM, Cuernavaca, México
<http://www.ccg.unam.mx/~vinuesa/>

Todo el material del curso (presentaciones, tutoriales y datos) lo encontrarás en:
<https://github.com/vinuesa/TIB-filoinfo>



• Tema 1 Introducción al Biocómputo en sistemas Linux

1. ¿Qué son la bioinformática y el biocómputo, y cómo pueden ayudarme para mi trabajo en biología?
2. ¿Qué es UNIX y Linux?
3. Tengo una PC que corre Windows, ¿cómo puedo correr Linux en mi máquina?
4. ¿Cómo hago trabajar a UNIX/Linux? – el shell: comandos y conceptos básicos
5. Sesión práctica: uso de tuberías de comandos UNIX/Linux para procesar archivos de secuencias



Introducción al biocómputo en sistemas UNIX/Linux

- ¿Qué es UNIX? - fuente: Wikipedia
- Unix (registrado oficialmente como **UNIX®**) es un sistema operativo portable, multitarea y multiusuario
- su desarrollo inicia en 1969 por un grupo de empleados de los **laboratorios Bell** de AT&T, entre los que figuran **Ken Thompson, Dennis Ritchie y Douglas McIlroy**
- UNIX es un Sistema Operativo no libre muy popular, porque está basado en una arquitectura que ha demostrado ser técnicamente estable a lo largo de sus >50 años de existencia.
- **MacOS X** es un derivado de UNIX BSD!

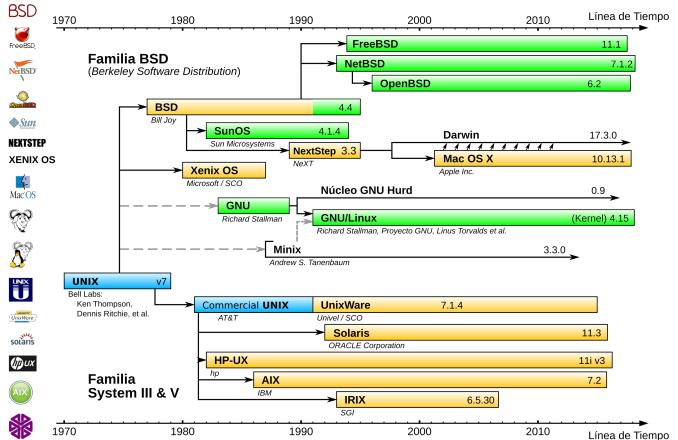


Ken Thompson y Dennis Ritchie
Fuente: Wikipedia

<https://www.fyccorp.com/articulo-unix:-la-simplicidad-del-ingenio>

Introducción al biocómputo en sistemas UNIX/Linux

- La evolución de las familias UNIX y clones - Fuente Wikipedia



Introducción al biocómputo en sistemas UNIX/Linux

- ¿Qué es Linux? - Evolución de sistemas UNIX y similares a UNIX (fuente: Wikipedia)

GNU: En 1983, Richard Stallman anunció el Proyecto GNU, un ambicioso esfuerzo para crear un sistema similar a Unix, que pudiese ser distribuido libremente. El software desarrollado por este proyecto -por ejemplo, GNU Emacs y GCC- también han sido parte fundamental de otros sistemas UNIX. (vean conferencias de R. Stallman en youtube ...)



Richard Stallman



Linus Torvalds

Linux: En 1991, cuando Linus Torvalds empezó a proponer el **núcleo Linux** y a reunir colaboradores, las herramientas GNU eran la elección perfecta. Al combinarle ambos elementos, conformaron la base del sistema operativo (basado en POSIX) que hoy se conoce como **GNU/Linux**.

Las **distribuciones** basadas en el núcleo, el software GNU y otros, como **Red Hat Linux** y **Debian GNU/Linux**, se han hecho populares tanto entre los aficionados a la computación como en el mundo empresarial y científico. Linux tiene un origen independiente, por lo que se considera un 'clón' de UNIX y no un UNIX en el sentido histórico.

Introducción al biocómputo en sistemas UNIX/Linux

- Tengo una PC que corre windows, ¿cómo puedo correr Linux en mi máquina?

1. Puedes instalar Linux en una nueva partición (Lo más recomendable).

descargas gratuitas de distribuciones desde:

=> **1. Ubuntu 22.04 LTS** - <http://www.ubuntu.com/getubuntu/download>

Ver instrucciones de instalación aquí:

<http://www.ubuntu.com/download/desktop/install-desktop-latest>

Centos - <https://www.centos.org/download/>

Fedora - <http://fedoraproject.org/es/get-fedora>

Biolinux - <http://environmentalomics.org/bio-linux/>

2. Puedes instalar MobaXterm, que proporciona una terminal para Windows con un servidor de ambiente gráfico X11, un cliente SSH para establecer sesiones remotas seguras con un servidor, diversas herramientas de red y más.

<https://mobaxterm.mobatek.net/download.html>

Introducción al biocómputo en sistemas UNIX/Linux

- Tipos de sesiones shell

1) local (acceder a tu computadora)

```
vinuesa@alisto:~/Cursos/TIB/TIB19-T3/sesion1_intro2linux$ ls
assembly_summary.txt.gz    linux_very_basic_commands_table.csv
intro_biocomputo_linux_pt1.odp Stenotrophomonas_complete_genomes_and_ftp_paths.txt
intro_biocomputo_linux_pt1.pdf working_with_linux_commands_code
linux_basic_commands.tab    working_with_linux_commands.html
linux_commands.tab        working_with_linux_commands.Rnd
vinuesa@alisto:~/Cursos/TIB/TIB19-T3/sesion1_intro2linux$
```

2) remota (login; conexión a otra máquina en Internet (chaac), vía protocolo ssh)

```
ssh -X usuario@ip; o ssh -X usuario@hostname
ssh -X vinuesa@132.***.***.18
```

```
Last login: Fri Aug 5 12:15:24 2022 from 132.248.220.40
Session lifetime based on X11 requested, but X11 support not compiled in.
(base) [vinuesa@buluc ~]$ hostname -I
132.248.220.16 10.1.1.10 192.168.122.1 172.17.0.1
(base) [vinuesa@buluc ~]$ hostname
buluc.lcg.unam.mx
(base) [vinuesa@buluc ~]$
```

Si tu máquina corre Windows, descarga el MobaXterm desde:

<https://mobaxterm.mobatek.net/download.html>

<https://mobaxterm.mobatek.net/download-home-edition.html>



Home Edition

Free

Full X server and SSH support
Remote desktop (RDP, VNC, Xdmcp)
Remote terminal (SSH, Telnet, rlogin, Mosh)
X11 Forwarding
Automatic GPGP browser
Master password protection
Plugins support
Portable and installer versions
Full documentation
Max. 12 sessions
Max. 2 SSH tunnels
Max. 4 macros
Max. 360 seconds for lftp, Nfs and Cron

Professional Edition

\$69 / 49€ per user*

* Excluding tax. Volume discounts available

Every feature from Home Edition +
Customize your startup message and logo
Modify your profile script
Remove unwanted games, screensaver or tools
Unlimited number of sessions
Unlimited number of tunnels and macros
Unlimited run time for network daemons
Enhanced security settings
12-months updates included
Deployment inside company
Lifetime right to use

En moodle encontrarás una detallada guía de instalación de mobaXterm

Sobre uso de la WiFi del auditorio Dr. Guillermo Soberón durante los TIB2024 en



EBM2022 Talleres Calendario Cuotas Registro Agenda

Pangenómica y Filogenómica Microbiana

Para un uso eficiente del ancho de banda limitado que tenemos para conectarnos al servidor en el que haremos todos los trabajos prácticos, pedimos de la manera más atenta, en el beneficio del buen funcionar del taller, que durante las sesiones de clases:

1. No conecten teléfonos u otros dispositivos móviles a la WiFi del auditorio y los pongan en silencio
2. Que no accedan a sus redes sociales o salgan con su browser a Internet desde su laptop
3. Desplieguen las instrucciones para las prácticas abriendo el archivo html correspondiente descargado en sus máquinas localmente desde el repositorio GitHub del taller

<https://github.com/vinuesa/TIB-filoinfo>

Introducción al biocómputo en sistemas UNIX/Linux

- explorando el nuevo ambiente ... El shell y la interpretación de comandos, unos ejemplos

1. – ¿Qué máquina es ésta a la que estoy conectado?

• **hostname** - read or set the hostname or the NIS domain name

```
[vinuesa@chaac] $ hostname
chaac.lcg.unam.mx
[vinuesa@chaac] $ hostname -i # corre también hostname --help
132.248.*.*
```

• **uname** - Print certain system information

```
[vinuesa@chaac] $ uname
Linux
[vinuesa@chaac] $ uname -a
Linux chaac.lcg.unam.mx 5.14.0-373.el9.x86_64 #1 SMP PREEMPT_DYNAMIC
Thu Oct 5 02:38:42 UTC 2023 x86_64 GNU/Linux
```

• **nproc** - Print the number of processing units available to the current process, which may be less than the number of online processors

```
[vinuesa@chaac] $ nproc
96
```

Licencia Creative Commons 4.0, no comercial
Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)
<https://creativecommons.org/licenses/by-nc/4.0/>

Introducción al biocómputo en sistemas UNIX/Linux

- explorando el nuevo ambiente ... El shell y la interpretación de comandos, unos ejemplos

1. – ¿Qué procesos se están corriendo en el sistema?

- **htop** - monitor system load

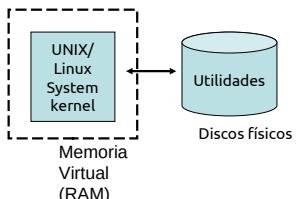
[vinuesa@chaac ~]\$ htop

Introducción al biocómputo en sistemas UNIX/Linux

- explorando el nuevo ambiente ... El shell (consola) y la interpretación de comandos

Los sistemas UNIX/Linux se dividen lógicamente en dos piezas: el **kernel** y las **utilidades**.

- ¿Qué es el shell?



El kernel es el corazón del sistema y reside en la memoria de la computadora desde el momento que se arranca y hasta que se apaga.

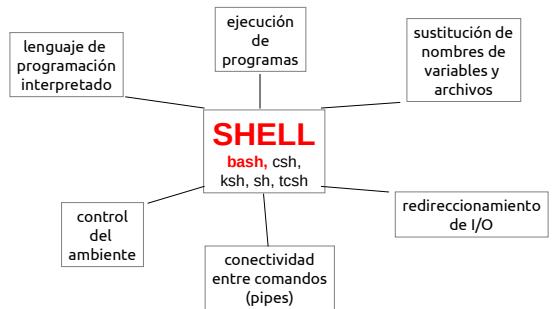
Las utilidades (comandos), residen en el **disco físico** y se cargan en memoria sólo cuando son llamadas.

El shell también es un programa. Se carga automáticamente en memoria desde que uno hace el login a una máquina para que el usuario pueda interactuar con ella, haciendo de interfaz entre el usuario el kernel y las utilidades.

Introducción al biocómputo en sistemas UNIX/Linux

- ## • explorando el nuevo ambiente ... El shell y la interpretación de comandos

- Las responsabilidades del shell -



Introducción al biocómputo en sistemas UNIX/Linux

- explorando el nuevo ambiente ... El shell y la interpretación de comandos

1. Mira estas entradas en Wikipedia:

http://en.wikibooks.org/wiki/Linux_Guide/Linux_commands

http://en.wikipedia.org/wiki/List_of_Unix_programs

2. Y corre estos comandos para ver parte de los programas del sistema

instalados en el servidor o en tu máquina:

19

ls /usr/bin

3. Un sencillo tutorial que todos deberían haber revisado ya (caps 1-5):

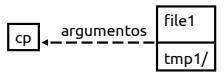
<http://www.ee.surrey.ac.uk/Teaching/Unix/>

Introducción al biocómputo en sistemas UNIX/Linux

- explorando el nuevo ambiente ... El shell y la interpretación de comandos
 - Ejecución de programas por el shell – ejemplos

- **Formato básico de un comando**

```
comando [argumento1 arg2 arg3 ...]  
-bash-3.1$  
-bash-3.1$ cp file1 tmp1/  
-bash-3.1$
```



¡los espacios separan argumentos!

necesito al menos 1 espacio entre comandos y argumentos!
Nombres de archivos en UNIX/Linux de preferencia no deben contener espacios

- Además de **argumentos**, los comandos pueden tener **opciones**, las cuales preceden a los argumentos y llevan un guión sencillo delante de una o más opciones.

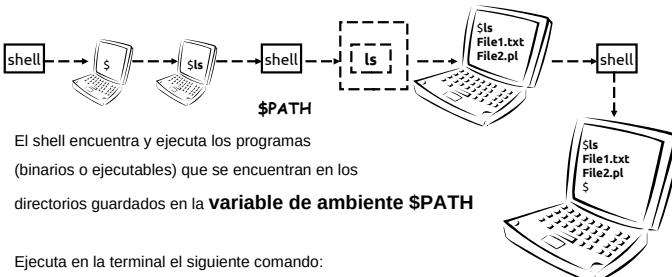
```
-bash-3.1$ ls -l /home # prueba también solo ls /home
drwxr-xr-x. 2 aagarcia students 4096 Oct 30 2014 aagarcia
drwxr-xr-x. 2 aanaya students 4096 Oct 30 2014 aanaya
drwxr-xr-x. 2 acampos students 4096 Oct 30 2014 acampos
drwxr-xr-x. 2 acarmona students 4096 Oct 30 2014 acarmona
drwxr-xr-x. 4 agodinez students 4096 Jul 1 2016 agodinez
drwxr-xr-x. 2 aguzman students 4096 Oct 30 2014 aguzman
```

Comandos Básicos

- ls (list)
 - \$ ls -l
 - \$ ls -a
 - \$ ls -la
 - \$ ls -l --sort=time # ls -ltr
 - \$ ls -l --sort=size -r #ls -lSr
 - \$ ls *.txt
 - cd (change directory)
 - \$ cd dir
 - \$ cd ./dir
 - \$ cd ..
 - \$ cd ../../
 - \$ cd /export/space2/tib/filo
 - pwd (print working directory)
 - \$ pwd
 - ~
 - \$ cd ~
 - ~user
 - \$ cd ~vinuesa
 - ¿Qué hará "cd ~vinuesa"?
 - ¿Y cd?
 - which
 - \$ which blastn
 - locate
 - \$ locate get_homologues.pl
 - \$ locate mi_archivo
 - find
 - \$ find . | grep stdio.h
 - \$ find /usr/include | grep stdio.h
 - \$ find . -type d
 - \$ find /home/vinuesa -name *TIB*
 - man (manual pages for command)
 - \$ man ls
 - \$ man find
 - \$ man man

Introducción al biocómputo en sistemas UNIX/Linux

- explorando el nuevo ambiente ... El shell y la interpretación de comandos
 - Tecleando comandos para el shell – el ciclo de comandos



Comandos Básicos (cont.)

- echo (print to STDOUT)
 - \$ echo "Hello World"
 - \$ echo -n "Hello World"
 - cat (concatenate)
 - \$ cat /proc/cpuinfo
 - \$ cat arch1 arch2
 - cp (copy)
 - \$ cp arch1 dir1
 - \$ cp -r dir1 ~vinuesa/tmp
 - mv (move or rename)
 - \$ mv arch1 archivo1
 - \$ mv arch1 ~vinuesa/tmp
 - mkdir (make directory)
 - \$ mkdir dir2
 - \$ mkdir -p dir2/practica1
 - rm (remove)
 - \$ rm arch1
 - \$ rm -rf dir2
 - less (paginador)
 - \$ less archivo.txt # q para salir
 - head (ver cabecera del archivo)
 - \$ head -5 archivo1.txt
 - tail (ver cola del archivo)
 - \$ tail -1 archivo.txt
 - \$ tail -f logfile.txt
 - sed (stream editor)
 - sed 's/esto/queello/' archivo.txt
 - vim (vi improved; a powerful command line text editor in Linux)
 - gedit (editor de texto con interfaz gráfica en gnome)
 - needit (otro editor de texto con interfaz gráfica)

Commandos Básicos (cont.)

- tar & gzip
 - \$ tar cvzf dir1.tar.gz dir1
 - \$ tar cvzf dir1.tgz dir1
 - \$tar cvpf dir1.tar dir1
- gzip (gnuzip, compress)
 - \$ gzip -9 dir1.tar
 - # genera dir1.tar.gz
- untar & ungzip
 - \$ tar xvfz dir1.tar.gz
- touch
 - \$ touch foo
- head
 - \$ head archivo.txt
- tail
 - \$ tail archivo.txt
 - \$ tail -f archivo2.txt
- |, >, grep, sort, cut, uniq
 - \$ ls -l /home | less
 - \$ cat /etc/passwd | cut -d: -f1 | sort
 - \$ ls -l /home | grep stud | wc -l > stud.txt
 - \$ ls -l /home | grep -v stud | sort | uniq -c
- backticks
 - \$ echo "The date is `date`"
 - \$ echo `seq 1 10`
- Hard, soft (symbolic) link
 - ln -s /path/to/gbk_files/*.gbk .
- scp (secure copy)
 - scp arch1 vinuesa@buluc.lcg.unam.mx: \$HOME/tmp

Commandos Básicos (cont)

- Uso del disco duro
 - \$ df -h /
- Uso disco de archivos
 - \$ du -sxh ~/
- **Uso avanzado: programación del Shell ☺**
 - **Asignación de variables y bucles for**
 - var1=123; echo \$var1; var2=/home/pepin && echo \$var2
 - for file in *faa; do muscle < \$file > \${file%.*}_aln.faa; done
 - for file in \$(ls *faa | grep rpoB); do echo -n \$file; grep -c '^' \$file; done

Linux text editors

- con entorno gráfico
 - **gedit**
 - nedit
- Sin entorno gráfico
 - **Vim**
 - emacs
 - nano
 - pico

Vim – the Linux power text editor

- 2 modes
 - Input mode
 - ESC to back to cmd mode
 - Command mode
 - Cursor movement
 - h (left), j (down), k (up), l (right)
 - ^f (page down)
 - ^b (page up)
 - ^ (first char.)
 - \$ (last char.)
 - G (bottom page)
 - :1 (goto first line)
 - Switch to input mode
 - a (append)
 - i (insert)
 - o (insert line after)
 - O (insert line before)
- Delete
 - dd (delete a line)
 - d10d (delete 10 lines)
 - d\$ (delete till end of line)
 - dG (delete till end of file)
 - x (current char.)
- Paste
 - p (paste after)
 - P (paste before)
- Undo
 - u
- Search
 - /
- Save/Quit
 - :w (write)
 - :q (quit)
 - :wq (write and quit)
 - :q! (give up changes)

Introducción al biocómputo en sistemas UNIX/Linux

- explorando el nuevo ambiente ... Trabajando eficientemente con el shell

¿Cómo me muevo en la línea de comandos?

- Usa **ctrl-e** para ir al final de la línea

- Usa **ctrl-a** para ir al principio de la línea

¿Cómo edito la línea de comandos?

- Usa la tecla **backspace** para eliminar uno a uno caracteres (del final hacia el principio)

- Usa **ctrl-w** para eliminar una palabra completa (del final hacia el principio)

- Usa **ctrl-u** para eliminar la línea completa (del final hacia el principio)

¿Cómo aborto o suspendo la ejecución de un comando?

Usa **ctrl-c** para abortar la ejecución del último comando

- Usa **ctrl-z** para suspender la ejecución del último comando

- Usa **bg** para poner este último comando a correr en el fondo (background)

Repetición de la ejecución de un comando y completado de nombres de comandos/archivos

- Unix recuerda los comandos ejecutados: usa flecha arriba o abajo para moverte por el historial de comandos "history" file o escribe **history | grep comando**

- Usa **TAB** para completar automáticamente el nombre de comandos, archivos o directorios

Introducción al biocómputo en sistemas UNIX/Linux

- explorando el nuevo ambiente ... Trabajando eficientemente con el shell

¿Cómo me muevo en la línea de comandos?

- Usa **ctrl-e** para ir al final de la línea

- Usa **ctrl-a** para ir al principio de la línea

¿Cómo edito la línea de comandos?

- Usa la tecla **backspace** para eliminar uno a uno caracteres (del final hacia el principio)

- Usa **ctrl-w** para eliminar una palabra completa (del final hacia el principio)

- Usa **ctrl-u** para eliminar la línea completa (del final hacia el principio)

¿Cómo aborto o suspendo la ejecución de un comando?

Usa **ctrl-c** para abortar la ejecución del último comando

- Usa **ctrl-z** para suspender la ejecución del último comando

- Usa **bg** para poner este último comando a correr en el fondo (background)

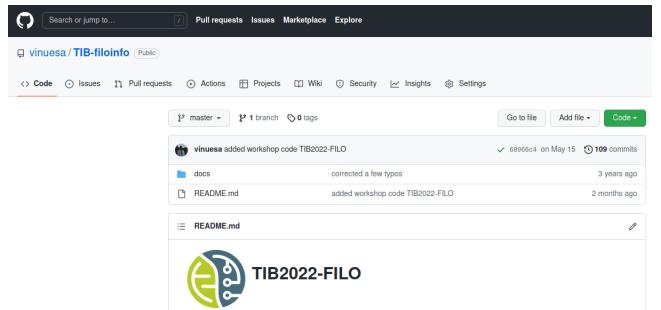
Repetición de la ejecución de un comando y completado de nombres de comandos/archivos

- Unix recuerda los comandos ejecutados: usa flecha arriba o abajo para moverte por el historial de comandos "history" file o escribe **history | grep comando**

- Usa **TAB** para completar automáticamente el nombre de comandos, archivos o directorios

Prácticas – asegúrate de tener clonado y actualizado el repositorio GitHub

<https://github.com/vinuesa/TIB-filoinfo>



1. Instala git en tu máquina (**sesión local**) con `$ sudo apt install git`
2. genera un directorio GitHub en tu \$HOME y clona el repositorio así:
`$ cd && mkdir GitHub && cd GitHub
git clone https://github.com/vinuesa/TIB-filoinfo.git`

Una vez clonado el repositorio, puedes abrir localmente los archivos html con las instrucciones para los ejercicios en tu navegador web de preferencia

Así evitamos consumir innecesariamente el limitado ancho de banda que nos proporciona la WiFi local. ¡Gracias por tu colaboración!



Sesión 1. Introducción al biocómputo en sistemas GNU/Linux

Pablo Vinuesa, Centro de Ciencias Genómicas - UNAM

2019-07-23

- 1 Sesión 1. Introducción al biocómputo en sistemas GNU/Linux - Primer contacto
 - 1.1 Conexión a un servidor y exploración de sus características básicas
 - 1.1.1 ssh establecer sesión remota encriptada (segura) vía ssh al servidor con número dado de IP
 - 1.1.2 hostname muestra el nombre del host (la máquina a la que estoy conectado) y la IP
 - 1.1.3 uname muestra el sistema operativo del host
 - 1.1.4 uptime muestra los procesos en ejecución y los recursos que consumen
 - 1.2 Exploración del sistema de archivos
 - 1.2.1 pwd imprime la ruta absoluta del directorio actual
 - 1.2.2 ls lista contenidos del directorio
 - 1.2.3 Expansión de caracteres con * y ?
 - 1.3 Movimientos por el sistema de archivos: comando cd
 - 1.3.1 cd de nuevo, ¿dónde estoy?
 - 1.3.2 cd .. cambia de directorio usando RUTA RELATIVA
 - 1.3.3 donde estoy?
 - 1.3.4 regresa a tu home
 - 1.3.5 cd cambiar directorios con rutas absolutas (/ruta/completa/al/dir) y relativas ../../